Replicator Formalism

A general account of replication

Ph.D. thesis

István Zachar
zac@caesar.elte.hu

Theoretical and Evolutionary Biology doctoral program,
Biology doctoral school,
Department of Plant Taxonomy and Ecology,
Eötvös Loránd University, Budapest

supervisor: Dr. Eörs Szathmáry, professor, CHAS
head of doctoral school: Dr. Anna Erdei, professor, MHAS
head of doctoral program: Dr. Eörs Szathmáry, professor, CHAS

2011 Budapest

Contents

Background

Evolution via natural selection is based on the successive, erroneous replication of informational molecules. A population of replicators is a distinguished set of (biologically) interesting entities capable of multiplication, and in some cases, evolution as well. Ever since the introduction of the term ‘replicator’ by Richard Dawkins in 1976, it spurred a great amount of discussion in various scientific. The replicator concept is useful and common enough to infiltrate into many fields of science, though it remains a question how different examples of various fields can be related and compared.

Informational replicators are responsible for complex adaptive systems, which renders their role paramount in evolutionary biology and related fields. Such highly successful adaptive systems include living organisms, the immune system, and perhaps cognitive capacities and language as well. The vague idea of memetics stems from the universal Darwinian concept of *multiplication*, *variability*, and *heredity*, being extended to culture. Studies of (self-)replication in chemical, (proto-)biological, linguistic, cultural, neuronal, digital, and kinetic systems have accumulated an impressive amount of theoretical and experimental knowledge in the last three decades. A multitude of replicator candidates and definitions are available in the literature. Examples include autocatalytic chemical cycles like the formose reaction, prions, viruses (even *in silico*), genes, genetic membranes, organelles and cells. In addition, there are many hypothetical systems, like ribozymes, hypercycles, the stochastic corrector model, the GARD model, the chemoton, and the neuronal replicator hypothesis. These existing or putative replicators, theoretically, all share identical or similar dynamics as they all are manifestations of the basic Darwinian principle.

Aim

The thesis is an undertaking to formalize and redefine the replicator concept, to both provide replicator-based studies with a firm background and to provide a modeling tool that could handle models of many present replication-selectional systems (*in vivo*, *vitro et silico*). Accordingly, the following questions were raised:

1. What is the qualitative and quantitative difference between genes, organisms, general autocatalytic intermediates and other multiplying entities?
2. What are the basic concepts behind the Mullerian criteria multiplication, heredity, and variability, and what part do they play in replication?
3. What is replicator identity and similarity? When do we say that two replicators are identical?
4. What is common and distinctive of genes and memes?

The problem of the replicator concept is its indefinite nature. Accordingly, five general, intuitive requirements were suggested that must be met by the reworked replicator definition:

1. **Autocatalysis**: potentially autocatalytic mode of generation.
2. **Similarity**: potentially above-chance similarity between parent and offspring.
3. **Informational**: ability to pass on information to offspring. This is a requirement for informational replicators.
4. **Inclusivity**: The widest definition has to include genes, memes, and possibly other entities as well.
5. **Specificity**: The definition must be able to distinguish between various multiplying entities: genes, memes, simple chemical cycle intermediates, lipid vesicles, kinetic multipliers, organisms, and higher-level entities in a clear and exact way.

To explore the consequences of the universality of replication and to cut through the bulk of misleading terms, this thesis introduces and details the Replicator Formalism, a generalized theory of autocatalytic entities. To better understand models and be able to relate results of various fields dealing with replication, the formalism applies a pure, objective language, independent of any specific field. It helps to recognize key features of autocatalytic systems, and it both provides a classification of replication systems, and suggests evolutionary trajectories for certain transitions from lesser systems to evolved ones.

Models and methods

The thesis focuses on the informational aspect of replication. The formalism is built on the principle that information might travel from entity-to entity, independent of material overlaps, during generative processes in replication. Accordingly, it is claimed that are three distinct and important inheritance methods available for replicating information in a selective environment: direct, Weismannian and Lamarckian methods. Three models were used to test various aspects of these inheritance methods.

First, an analytic model is presented to estimate the probability of replicating a sequence correctly in case of Lamarckian (or any other) inheritance. The model assumes simple sequences for the various life-stages of the replicator, like DNA and protein sequences. During replication, these sequences transform to each other. It is possible to estimate the probability to end up with the same sequence from where replication started. An analytical approach was used to calculate estimates. Results were checked against 1000 iterations of explicit simulations of replication processes.

Second, a stochastic model is built using the background formalism, to test a possible scenario for the emergence of division of labour. The model assumes a simple, directly replicating, informational, autocatalytic template molecule, which produces a neutral side product. According to the model, the side product is also informational, and might provide positive contribution to fitness and/or to the heritable genotype. Sequences and genotype-phenotype-specifications are explicitly modelled, and inherited during replication. It is expected that if the appropriate fitness landscape (*Nk*-type correlated) is selected and the transmission fidelity of various channels are set correctly, the two entities separate roles, and a division of labour emerges ultimately, where one entity is responsible for storing heritable information stably, while the other is responsible to providing a good-fitnessed phenotype.

A third model was used to test the dynamic ecology of multiple informational replicators in a reproducing vehicle: the chemoton. A deterministic model was built to numerically test the possible coexistence of informational replicators in a growing and splitting protocell. Results suggest a possible solution to overcome the problem of prebiotics: how to integrate information without implying hard competition between entities (or large error rate for large templates).

Theses

The abstract framework of the Replicator Formalism is presented. It allows the proper identification of various types of replicators. It helps to understand the old problem of the units and levels of selection and evolution. Using the formalism as a conceptual and as a modeling tool, and based on the models detailed above, the following results and conclusions were drawn:

1. The formalism provides a clear and logical way to identify, represent, and classify replicators. Several concepts (like genotype, interactor, vehicle, translation, etc.) were integrated into the formalism, to give a full account of the universal principle of replication.
2. The formalism gives consistent definitions for replication, informational replication, and subtypes of informational replication (like Weismannian and Lamarckian inheritance).
3. A classification for the partition of the replicator-continuum is provided based on the multiplication, variability and heritability criteria, where specific replicators are nested within ones that are more general. The classification can be applied to known replicators and to future candidates.
4. Further dimensions of the replicator space are identified and integrated into the formalism and classification to give the widest possible coverage of various fields.
5. Replicator equivalency is a matter of relativity, and can only be defined from an arbitrary point of view. A formal definition is provided for replicator equivalency.
6. Heterocatalytic side products of autocatalytic cycles are replicators as well. This fits proteins expressed from genes, but also nests of birds. Such products, providing heterocatalytic aid to the parent cycle, ultimately help themselves.
7. The difference between genes and organisms is claimed to be quantitative at the first place. Accordingly, all such reproducing entities (like DNA sequences, cells, organism, etc.) are considered replicators. A model is presented to distinguish between such replicators based on the amount of variability they can acquire and inherit.
8. There could be possibly different material, informational and selectional topologies behind inheritance systems.
9. Phenotypic replication is direct replication lacking a distinct interactor.
10. Informational replication is about the replication of a non-material pattern, thus material overlap is not required for replication at all (as a general criterion).

11. There is a continuum between direct and Weismannian replication, Lamarckian being in-between them.

12. This continuum can be traversed by evolution: an evolutionary scenario for the emergence of division of labour is suggested, and tested in a stochastic model.

13. The loss of information in replication can only be understood in case of successive elementary transformations. Accordingly, information loss can originate from 1) exogenous mutation, 2) probabilistic translation, 3) real lossy transmission.

14. Information can be stably maintained in Lamarckian systems. This is supported by both analytical estimations and stochastic simulations.

15. Lamarckian replication is stable enough to compete with Weismannian inheritance. Results of a simple stochastic model support this.

16. Informational replicators can coexist in a vehicle even if there is only stoichiometric coupling, but no explicit phenotypic selection. Numerical simulations of the chemoton model with multiple competing templates suggest a stable coexistence and a resolution of the paradox of prebiotics.

Conclusions

The Replicator Formalism is presented in the thesis. It gives a bottom-up theoretical approach to redefine the concept of the replicator. The new, formal definition and the abstract models presented can distinguish between possible replicator types, based on their quantity of variable and heritable information. This allows for the exact identification of various replicator types and their underlying dynamics.

Preliminary models provide an evolutionary scenario for the appearance of different informational topologies, and a method to integrate informational replicators in a vehicle, against competition. The Replicator Formalism gives a powerful theoretical and modeling framework, providing a tool to bridge various models and results of different fields.

Publications

Eörs Szathmáry, Zoltán Szatmáry, Péter Ittzés, Gergő Orbán, István Zachar, Ferenc Huszár, Anna Fedor, Máté Varga, and Szabolcs Számadó, 2006. In silico *evolutionary developmental neurobiology and the origin of natural language*, in: Lyon, C. and Nehaniv, C. L. and Cangelosi, A. (eds.): Emergence of communication and language, chapter 8, pages 151–184. Springer-Verlag New York, Inc., New York.

István Zachar and Eörs Szathmáry, 2010. A New Replicator: A theoretical framework for analyzing replication. *BMC Biology*, 8(21): 21 pp.

István Zachar, 2011. The feasibility of segmentation of protolanguage. *Interaction Studies*, 12(1): 1–35.

István Zachar, Ádám Kun, Chrisantha Fernando, and Eörs Szathmáry, 2010. *Replicators. From molecules to organism*. Pan Stanford Publishing, *accepted.*

István Zachar, Anna Fedor, and Eörs Szathmáry, 2011. Two different template replicators coexisting in the same protocell: Stochastic simulation of an extended chemoton model. *PLoS ONE. Accepted*.